

An Iterated Cochrane-Orcutt Procedure for Nonparametric Regression

by

Naomi S. Altman

BU-1026-MA

April 1991

An Iterated Cochrane-Orcutt Procedure for Nonparametric Regression

Naomi Simone Altman

Biometrics Unit

Cornell University

This work was supported by Hatch Grant 151410 NYF.

Abstract

Altman (1990) and Hart (1991) have shown that kernel regression can be an effective method for estimating an unknown mean function when the errors are correlated. However, the optimal bandwidth for kernel smoothing depends strongly on the correlation function, as do confidence bands for the regression curve.

In this paper, the simultaneous estimation of the regression and correlation functions is explored. An iterative technique analogous to the iterated Cochrane-Orcutt method for linear regression (Cochrane and Orcutt, 1949) is shown to perform well. However, for moderate sample sizes, stopping after the first iteration produces better results.

An interesting feature of the simultaneous method is that it performs best when different kernels are used to estimate the regression and correlation functions. For the regression function, unimodal kernels are known to be optimal. However, examination of the mean squared error of the correlation estimator suggests that a bimodal kernel will perform better for estimating correlations. Use of a bimodal kernel for estimating the correlation function, followed by use of a unimodal kernel for estimating the regression function at the final step, performed best in a simulation study.

Keywords: autocorrelation; smoothing; nonlinear time series

1. Introduction

Altman (1990) and Hart (1991) have shown that kernel regression can be an effective method for estimating an unknown mean function, $\mu(x)$ in the nonparametric regression problem

$$y = \mu(x) + \varepsilon$$

when the errors, ε , come from a stationary correlated process with summable correlations. The optimal bandwidth for kernel smoothing depends strongly on the correlation function. Altman (1990), Hart (1991) and Truong (1989) have suggested methods for estimating the correlation function from the data. This paper explores the use of iterative, Cochrane-Orcutt (CO) type techniques, for simultaneous estimation of the regression and correlation functions.

The bias and variance of both the regression and correlation estimates are controlled by the kernel and bandwidth used for smoothing. If the bandwidth needs to be chosen from the data, as is usually the case, estimation of the correlation function is critical.

Altman (1990) and Hart (1991) show that cross-validation (CV) and related techniques for estimating prediction error perform poorly when the errors are correlated. Their mean squared error (MSE) computations, which are summarized in Section 2, indicate that “plug-in” methods (Gasser, Kneip, and Köhler, 1991) may also need to be adjusted for correlation. Overestimation of the correlations leads to estimated bandwidths which are too large, and hence to oversmoothing the data, while underestimation leads to undersmoothing. Two methods for adjusting CV have been suggested (Altman, 1990; Chiu, 1989; Diggle and Hutchinson, 1989; Engle et al, 1986; and Hart, 1991) and are reviewed in Section 3.

The asymptotic properties of the MMEs of correlation are reviewed in Section 2. Over most reasonable ranges of bandwidths, oversmoothing leads to positive bias in the method of moments estimators (MMEs) of correlation, while undersmoothing leads to negative bias. Thus, using iterative methods, it is possible to enter a positive feedback

loop in which both the regression and correlation estimators move away from their true values as iteration continues.

At the asymptotically optimal bandwidth, λ^* , for estimating the regression function, the asymptotic MSE of the MMEs of correlation are proportional to a characteristic function of the kernel, and depend, to first order, only on the bias. For unimodal kernels, the asymptotic bias at λ^* is always negative. However, the asymptotic bias at λ^* can be reduced to a lower order of magnitude by an appropriately chosen bimodal kernel. This suggests that certain bimodal kernels may be preferable to unimodal kernels for use in CO type computations.

By contrast, Gasser and Müller (1979) show that optimal kernels for estimating the regression function are unimodal when the errors are independently and identically distributed (i.i.d.). Altman (1990) showed that this result is also true when the errors are correlated. Therefore, this paper proposes that the final step in the CO procedure use a unimodal kernel to estimate the regression function.

Section 4 is a simulation study, supporting the use of the bimodal kernel and the iterative method. Section 5 is an example of the use of the method to determine features in a long time series of daily sea surface temperatures.

2. Asymptotic MSE of the Regression and Correlation Estimators

This paper focuses on the nonparametric regression model,

$$y_{n,i} = \mu\left(\frac{i}{n}\right) + \varepsilon_i \quad (1)$$

where n is the sample size, $\mu(x)$ is a smooth deterministic mean function on $[0, 1]$, and ε is a stationary second order error process with mean zero and covariance function

$$\text{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \rho(|i - j|). \quad (2)$$

In this formulation, the design points become closer together as the sample size increases, but the error process remains the same. This model is discussed by Altman (1990) and

Hart (1991). The model is of practical application in situations in which the correlation is induced by the measuring device, for example, when the output from a monitor is a filtered sequence. In many applications, the correlation function depends on the distance between design points, and thus on n . However, for these applications, asymptotic results based on model (1) can be viewed as approximations valid for large sample sizes.

Under regularity conditions on the mean and correlation functions, kernel regression has been shown to be consistent for estimating the mean function for classes of models which include (1) (Altman, 1990 and Hart 1991). It is therefore natural to attempt to estimate functionals of the error distribution from the residuals

$$e_{\lambda,n}(i) = y_{n,i} - \hat{\mu}_{\lambda,n}\left(\frac{i}{n}\right). \quad (3)$$

The regression estimator, $\hat{\mu}_{\lambda,n}\left(\frac{i}{n}\right)$ is the kernel estimator of Priestley and Chao (1972). This has the form

$$\hat{\mu}_{\lambda,n}(x) = \sum_{j=0}^n \frac{K\left(\frac{x - \frac{j}{n}}{\lambda}\right)}{n\lambda} y_{n,j}. \quad (4)$$

K is called the kernel function, and λ is a smoothing parameter, called the bandwidth.

Only kernels with the following properties are considered:

- A) K is symmetric about 0.
- B) K has support only on the interval $(-\frac{1}{2}, \frac{1}{2})$.
- C) K is Lipschitz continuous of order $\alpha > 0$.

K is called a kernel of order p if all the first $p - 1$ moments of K are 0, and the p^{th} moment,

$$s_K = \int x^p K(x) dx \quad (5)$$

is not zero. The squared norm of K

$$W_K = \int K^2(x) dx \quad (6)$$

is also needed for the computations that follow.

Altman (1990) showed that kernel estimators of this type are consistent estimators of the mean function under the following regularity conditions:

D) The mean function μ has square integrable p^{th} derivative ($p \geq 2$) which is Lipschitz of order γ , $0 < \gamma \leq 1$.

E) The correlation function is absolutely summable and

$$\sum_{j=1}^{\infty} \rho(j) = S_{\rho}. \quad (7)$$

F) The correlation function satisfies:

$$\frac{1}{N} \sum_{i=1}^N i |\rho(i)| = o(1)$$

Conditions E and F are mixing conditions which ensure that observations sufficiently far apart are essentially uncorrelated.

Theorem 1: Suppose the kernel satisfies A-C and is of order p , the regression function satisfies D for the same p , and the correlations satisfy E and F. Then, for the kernel estimator with bandwidth λ , the asymptotic mean integrated squared error (AMISE) of the regression function over the region $\lambda/2 \leq x \leq 1 - \lambda/2$ is

$$\begin{aligned} AMISE &= \lim_{n \rightarrow \infty} E \left(\int_{\lambda/2}^{1-\lambda/2} (\hat{\mu}_{\lambda,n}(x) - \mu(x))^2 dx \right) \\ &= \sigma^2 \left(\left(\frac{\lambda^p s_K}{p!} \right)^2 N(\lambda) + \frac{W_K}{n\lambda} (1 + 2S_{\rho}) \right) + o\left(\frac{1}{n\lambda}\right) + o(\lambda^{2p}). \end{aligned}$$

The signal to noise ratio, $N(\lambda)$, is defined by:

$$N(\lambda) = \frac{\int_{\frac{\lambda}{2}}^{1-\frac{\lambda}{2}} (\mu^{(p)}(x))^2 dx}{\sigma^2}. \quad (8)$$

The proof of Theorem 1 is in Altman, 1990.

Corollary (1): Under the conditions of Theorem 1, the asymptotically optimal bandwidth, λ^* is

$$\lambda^* = \left[\frac{W_K(1 + 2S_\rho)}{pN(0) \left(\frac{p!}{s_K} \right)^2} \right]^{\frac{1}{2p+1}} n^{-\frac{1}{2p+1}}. \quad (9)$$

For the consistency of the MMEs of variance and correlation based on the residuals from the kernel smooth, further regularity conditions on the errors are needed. A condition which is often used in time series analysis (see, for example, Brockwell and Davis, 1987, chap. 7) and which is sufficient for consistency is:

$$\begin{aligned} \text{G) } \varepsilon_{n,t} &= \sum_{j=-\infty}^{\infty} \psi_j z_{t-j} && \text{with } \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \\ z_t &\text{ i.i.d.} && \text{with } E(z_t) = 0 \\ &&& E(z_t^2) = \sigma^2 \\ &&& E(z_t^4) < \infty. \end{aligned} \quad (10)$$

For results on the variance of the MMEs, the following condition is needed on the fourth moment of the error process:

$$\begin{aligned} \text{H) For all } r, s, t \\ \sum_{n=0}^{\infty} E(\varepsilon_t \varepsilon_{t+s} \varepsilon_{t+r} \varepsilon_{t+r+s+n}) &\text{ converges.} \\ \sum_{n=0}^{\infty} E(\varepsilon_t \varepsilon_{t+s} \varepsilon_{t+r} \varepsilon_{t+r-s+n}) &\text{ converges.} \end{aligned}$$

Define

$$\hat{\gamma}_{\lambda,n}(s) = \frac{1}{n} \sum_{i=\lceil \frac{n\lambda}{2} \rceil}^{n-\lceil \frac{n\lambda}{2} \rceil - s} e_{\lambda,n}(i) e_{\lambda,n}(i+s) \quad (11)$$

where $e_{\lambda,n}(i) = y_{n,i} - \hat{\mu}_{\lambda,n}(\frac{i}{n})$.

Theorem 2: Suppose the data and kernel satisfy conditions A - H and define $\hat{\gamma}_{\lambda,n}(s)$ by (11). For fixed s , n and λ , define the method of moments estimator of $\rho(s)$ by

$$\hat{\rho}_{\lambda,n}(s) = \frac{\hat{\gamma}_{\lambda,n}(s)}{\hat{\gamma}_{\lambda,n}(0)}. \quad (12)$$

Then $\hat{\rho}_{\lambda,n}(s)$ is asymptotically normal and, as $\lambda \rightarrow 0$ and $n\lambda \rightarrow \infty$,

$$E(\hat{\rho}_{\lambda,n}(s)) = \frac{\rho(s) + \lambda^{2p} \left(\frac{s_K}{p!}\right)^2 N(\lambda) + \frac{(1+2S_\rho)}{n\lambda}(W_K - 2K(0))}{1 + \lambda^{2p} \left(\frac{s_K}{p!}\right)^2 N(\lambda) + \frac{(1+2S_\rho)}{n\lambda}(W_K - 2K(0))} + o(\lambda^{2p}) + o\left(\frac{1}{n\lambda}\right) + o\left(\frac{s}{n}\right). \quad (13)$$

and

$$Var(\hat{\rho}_{\lambda,n}(s)) = \frac{V_s}{n} + o\left(\frac{1}{n^2\lambda^2}\right)$$

where V_s is given by Bartlett's formula (Bartlett, 1946) for the process ε_t .

The proof of Theorem 2 is in Altman (1991).

Theorem 2 shows the consistency of the MMEs of the correlation function. It also shows that, asymptotically, the mean squared error of the correlation estimates is dominated by the bias, which depends on the noise to signal ratio $N(\lambda)$ and the sum of the correlations S_ρ .

While the MMEs are consistent for all lags, it is important to note that the coefficients of the higher order terms increase in magnitude with s . In finite samples, correlations at low lags are estimated much more precisely than correlations at longer lags.

Corollary 2, below, discusses the dependence of the estimation bias on $N(\lambda)$ and S_ρ .

Corollary 2: Under the conditions of Theorem 2, asymptotically,

- a) If $W_K \geq 2K(0)$, the bias of $\hat{\rho}_{\lambda,n}(s)$ is positive.
- b) If $W_K < 2K(0)$, $\hat{\rho}_{\lambda,n}(s)$ has bias which is increasing in λ , and the signal to noise ratio, $N(\lambda)$, and is decreasing in S_ρ .

Proof: Let $r_{\lambda,n}(s) = (\rho(s) + C_{\lambda,n}) / (1 + C_{\lambda,n})$ where

$$C_{\lambda,n} = \lambda^{2p} \left(\frac{s_K}{p!}\right)^2 N(\lambda) + \frac{(1+2S_\rho)}{n\lambda}(W_K - 2K(0)). \quad (14)$$

Then

$$E(\hat{\rho}_{\lambda,n}(s)) = r_{\lambda,n}(s) + o(\lambda^{2p}) + o\left(\frac{1}{n\lambda}\right). \quad (15)$$

$r_{\lambda,n}(s)$ is a hyperbola in $C_{\lambda,n}$, with asymptote 1, and singularity at $C_{\lambda,n} = -1$. $r_{\lambda,n}(s)$ is increasing in $C_{\lambda,n}$ for $C_{\lambda,n} > -1$. If $W_K \geq 2K(0)$, then $C_{\lambda,n} > 0$. If $W_K < 2K(0)$, then, for λ sufficiently small, and $n\lambda$ sufficiently large, $C_{\lambda,n} > -1$ and $r_{\lambda,n}(s)$ is increasing in λ and $N(\lambda)$ and decreasing in S_ρ .

Computing $C_{\lambda^*,n}$ gives

$$C_{\lambda^*,n} = \frac{(1 + 2S_\rho)}{n\lambda^*} (W_K(1 + \frac{1}{2p}) - 2K(0)) \quad (16)$$

This suggests that a kernel of order p will perform well in a simultaneous estimation scheme if $W_K = \frac{4p}{2p+1}K(0)$. Such a kernel will be denoted a CO kernel. Suitable kernels can be computed using truncated polynomials.

Altman (1990) suggested a two-step estimation procedure for estimating the mean and correlation functions. Residuals from a moderate bandwidth smooth were used to compute MMEs of the low order correlations. These were extended to all lags by assuming an autoregressive moving average (ARMA) model for the errors. These estimates were used with a corrected bandwidth selection technique for estimating the regression function. Simulations using the same unimodal kernel to estimate the regression and correlation functions showed that this method worked well for estimating the mean function, but less well for estimating the correlations.

Improvement of this technique by iteration has been suggested (Bates, 1985; Hart, 1991). Corollary 2 suggests that a CO kernel should perform better than an unimodal kernel in an iterative scheme.

There is no guarantee that a CO kernel will be optimal for estimating either the regression or correlation function. In fact, since a CO kernel is bimodal, it cannot be optimal for estimating the regression function. However, for the unimodal kernels which are optimal for estimating the regression function, $K(0) \geq W_K$, so that $C_{\lambda^*,n}$ is negative, and the correlation estimates tend to be biased down. The simulation results presented in Section 4 confirm that estimation procedure performs best when a CO kernel is used to estimate the correlation function, and a unimodal kernel is used at the final step to

estimate the regression function.

3. Correcting Cross-Validation for Correlation

Altman (1990) and Hart (1991) showed that CV and related bandwidth selection techniques perform very poorly when the errors are correlated, unless appropriate corrections are made. The problem is that the uncorrected bandwidth selectors, which may be regarded as (almost) unbiased estimators of squared prediction error in the case of i.i.d. errors, are highly biased when the errors are correlated. Two corrections have been suggested when the correlation is known, and are discussed in Altman (1990).

The indirect correction can be viewed as a likelihood based approach (see, for example, Diggle and Hutchinson, 1989, and Engle et al, 1986). The idea is to transform the residuals using the inverse square root of the correlation matrix. Ordinary cross-validation is then done using the transformed residuals.

The direct correction is simply a bias correction, to make the cross-validation criterion asymptotically unbiased for squared prediction error:

$$CV_\rho = \sum_{i=1}^n \frac{e_{\lambda,n}^2(i)}{(1 - \sum_{j=1}^n w_{\lambda,n,i}(\frac{j}{n})\rho(i-j))^2}.$$

Similar corrections can be made to generalized CV, and other criteria which estimate squared prediction error. The direct correction is used in the studies discussed in this paper. However, all analyses were also carried out using the indirect correction, with very little change in the results.

Analysis of the bandwidth selected by either criterion shows that, when the correction is based on the true correlation function, the selected bandwidth converges to the asymptotically optimal bandwidth. If, however, the estimated correlations used for correction are too big (in the sense that S_ρ is too big) the selected bandwidth will tend to be too large, while, if the estimated correlations are too small, the converse is true.

4. Simulation Results

A simulation study was carried out to test the suggestions of Section 2. 128 points were generated from the process

$$y_{n,i} = \cos(3.15\pi i/n) + \varepsilon_i$$

where the error process was AR(1). Various values of $\rho(1)$ and two variances, $\sigma^2 = 1.0$ and $\sigma^2 = 0.01$ were used. 50 samples were generated for each combination of variance and correlation.

For each experiment, 4 second order kernels and a fourth order kernel were used. The definitions of the kernels are displayed in Table 1. (The kernels are zero on $(-\frac{1}{2}, \frac{1}{2})^c$.) The fourth order “spline” kernel is a truncated version of the effective kernel for the cubic smoothing spline with equally spaced design points (Silverman, 1984 and 1985).

Table 1. Kernels

Kernel	equation	$K(0)$	W_K
uniform	1	1	1
triangular	$4(\frac{1}{2} - x)$	2	1.333
quadratic	$6(\frac{1}{4} - x^2)$	1.5	1.2
spline	$8\exp(- x /16\sqrt{2})\sin(x /16\sqrt{2} + \pi/4)$	$4\sqrt{2}$	$3\sqrt{2}$
bimodal	$-74.876t^2(t^2 - .21987) + .56405$.56405	.90248

The bimodal kernel is a quartic second order kernel chosen to be symmetric about zero, vanish at the points $-1/2$ and $1/2$ and to be a CO kernel.

Bandwidths for the spline kernel were multiplied by 3, as preliminary investigations showed empirically that, for the sample sizes and mean function used in this study, and i.i.d. errors, the optimal bandwidth for this kernel is about three times as large as for the others. The MME of the first autocorrelation was based on the same design points for all the kernels, so the estimates for the spline kernel include some endpoint effects.

In the initial simulations, the starting bandwidth was chosen by minimizing the true Total Squared Error (TSE) on a grid of values. Residuals from this smooth were then

used to obtain the initial estimate of $\rho(1)$, which was subsequently used to correct CV for bandwidth selection. Thus iteration was started from a bandwidth very near the optimal one for the realization. Iteration was then done until “convergence”. The computation was deemed to have converged if either $\hat{\rho}_{n,\lambda}(1)$ converged, or the iterations cycled. Cycling occurred when, for two bandwidths λ_1 and λ_2 , CV was minimized for $\hat{\rho}_{n,\lambda_1}(1)$ at λ_2 and for $\hat{\rho}_{n,\lambda_2}(1)$ at λ_1 . When this happened, the final estimate of correlation was found by averaging $\hat{\rho}_{n,\lambda_1}(1)$ and $\hat{\rho}_{n,\lambda_2}(1)$

For the initial simulations, the same kernel was used to estimate the correlation and regression functions. In the final set of simulations, the bimodal kernel was used to estimate the correlation function, and the quadratic kernel was then used to estimate the regression function.

When analyzing real data, identifying a suitable bandwidth to start iteration is problematic. The simulations were redone, starting at a moderate ($\lambda = .15$) bandwidth, instead of at λ^* . However, this made little difference to the value of the bandwidth at convergence. Results are not reported here.

All the kernels gave reasonable estimates of $\rho(1)$ for both values of the variance, when the true value was negative. For positive $\rho(1)$, the estimates were biased down, and were particularly poor when the variance was small. The spline kernel did very poorly, often failing to converge for positive $\rho(1)$.

The bimodal kernel was best for estimating the correlation when $\rho(1)$ was positive. Of the other kernels, only the uniform kernel, with the smallest value of $W_K(1+1/2p)-2K(0)$, performed well for positive $\rho(1)$. However, it did not perform as well as the bimodal kernel. The estimates of $\rho(1)$ from the bimodal and uniform kernels are displayed in Figure 1. The bimodal kernel was designed to reduce bias, but it also appears to have reduced variance, particularly at $\rho(1) = .9$.

The bimodal kernel is poor for estimating the regression function. Figure 2 displays the ratio, for each data set, of minimum TSE achieved using the bimodal kernel over the minimum TSE achieved using the quadratic kernel. The quadratic kernel was uniformly

better except for large positive $\rho(1)$.

When the bimodal kernel is used to estimate the correlation, and the quadratic kernel is used to estimate the regression function, the iterative procedure performs very well. Figure 3 displays the ratio, for each data set, of the TSE achieved using the iterative procedure, over the minimum TSE achieved using the quadratic kernel. The ratio can be as large as 10 when the variance is large, but is generally less than 2. By contrast, when the correlation function is also estimated using the quadratic kernel, CV selects regression estimators with TSE 20 to 40 times greater than the minimum.

In linear least squares regression, it has been noted that estimating the error correlation and using weighted least squares does not improve the TSE unless the sample size or correlation is very large, due to the additional variance introduced by estimating the weights. In kernel regression, when the bandwidth is selected using an estimator of squared prediction error, estimating the correlation is crucial (Altman, 1990 and Hart, 1991). It is useful to ask, however, whether the iterative procedure provides any improvement over a two-step procedure. To determine this, estimation was started with bandwidth .25, instead of the optimal bandwidth. The two-step and fully iterative procedures were then compared. The results are displayed in Figure 4.

It is notable that, for all combinations of parameters, the two-step procedure outperformed the full iterative procedure.

To determine the effect of sample size, the simulation study was repeated with sample size 1024, and appropriated adjusted initial bandwidth of .16. For $\sigma^2 = 1.0$, the iterative procedure marginally outperformed the two-step procedure. For $\sigma^2 = 0.01$, the two-step procedure continued to outperform the iterative procedure.

5. Example

The two-step procedure was used to estimate a suitable correlation and mean function for a set of 4380 sea surface temperatures collected daily at Granite Canyon, California (Breaker, Lewis and Orav, 1984). This data set is known to have asymmetric seasonal

and other periodic effects; however, the features of most interest are broad peaks, called El Niño episodes, associated with unusual weather patterns in the middle latitudes. This data was analyzed in Altman, (1990), using a two-step procedure, with the same unimodal kernel used in both steps. The analysis was done assuming that the errors came from an AR(2) process. A purely nonparametric approach was used for estimating the regression function, with no attempt to model the seasonal effects. Using the quadratic kernel, the first two autocorrelations were estimated to be 0.85 and 0.69.

In the current analysis, the bimodal kernel was used, with an initial bandwidth of a quarter of a year, to generate residuals for estimating the autocorrelations. The AIC criterion (Akaike, 1970 and 1974) was used to determine the degree of a suitable AR process for estimating the autocorrelations. The estimated autocorrelations were then used with the quadratic kernel to estimate the mean function.

The degree chosen by AIC was 2. (AIC was also used with the residuals from the quadratic smooth, to confirm that the earlier analysis was appropriate.) The first two estimated autocorrelations were 0.86 and 0.72. The bandwidth chosen by the corrected CV criterion was 124 days, compared to 95 days in the earlier study. However, the final curve estimates were quite similar, never differing by as much as 0.4°C (compared to residuals as large as 4.0°C).

6. Conclusions

When the correlations are sufficiently short-term, kernel regression estimators converge at the same rate, (although with a different constant), as when the errors are i.i.d. Convergence relies, however, on choice of an appropriate sequence of bandwidths.

Automatic selection of bandwidth from the data is often desirable. Cross-validation and related techniques need to be adjusted for correlation. Appropriate confidence bands for the kernel regression estimate also depend on the correlation structure.

For these reasons, even if estimating the regression function is the primary goal, it is desirable to estimate the correlation function from the data. MMEs based on residuals

from the smooth are consistent, but may be very biased in finite samples unless the kernel function and bandwidth are chosen carefully.

For unimodal kernels, oversmoothing leads to estimates of the correlation which are biased up, and undersmoothing to estimates which are biased down. In turn, corrected bandwidth selection techniques pick the bandwidth too large when the correlation is overestimated, and too small when it is underestimated. The result is that iterative techniques tend to do poorly when unimodal kernels are used.

If appropriate bimodal kernels are used for estimating the correlation function, iterative techniques for simultaneous estimation of the mean and correlation functions have been shown to perform well. However, for moderate sample sizes, a two-step procedure, in which first the correlations and then the regression function are estimated, outperforms the iterative procedure.

References

- Akaike, H. (1970) "Statistical predictor information," *Annals of Statistical Mathematics* 2 203-217.
- Akaike, H. (1974) "A new look at the statistical model identification," *IEEE Transactions of Automatic Control* 19 716-723.
- Altman, N.S. (1990) "Kernel Smoothing of Data with Correlated Errors," *Journal of the American Statistical Association* 85 749-759.
- Altman, N.S. (1991) "Estimation of Error Correlation in Kernel Regression," Memo 982-M, Cornell University, Biometrics Unit.
- Bates, P.J. (1985) Discussion of "Some aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting," by B.W. Silverman, *Journal of the Royal Statistical Society Series B* 47 1-52.
- Bartlett, M.S. (1946) "On the theoretical specification and sampling properties of auto-correlated time series," *Supplement of the Journal of the Royal Statistical Society* 8 27-41.

- Box, G.E.P. and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*, Holden-Day, Inc. San Francisco.
- Breaker, L.C. and Lewis P.A.W. (1988) "A 40-50 Day Oscillation in Sea-Surface Temperature Along the Central California Coast," *Estuarine Coast and Shelf Science* 26 395-408.
- Brockwell, P.J. and Davis, R.A. (1987) *Time Series: Theory and Methods*, Springer-Verlag, New York.
- Chiu, S. (1990) "Bandwidth Selection for Kernel Estimate with Correlated Noise," *Statistics & Probability Letters* 8 347-354.
- Cochrane, D. and Orcutt, G.W. (1949) "Application of Least Square Regressions to Relationships Containing Auto-correlated Error Terms," *Journal of the American Statistical Association* 44 32-61.
- Diggle, P.J. and Hutchison, M.F. (1985) "On Spline Smoothing with Autocorrelated Errors," *Australian Journal of Statistics* 31 166-182.
- Engle, R. Granger C.W.J. Rice, J. Weiss A. (1986) "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association* 81 310-320.
- Gasser, T., Kneip, A., Köhler, W. (1991) "A Flexible and Fast Method for Automatic Smoothing," (to appear in the *Journal of the American Statistical Association*).
- Gasser, T., Müller, H-G. (1979) "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation* 23-37, Lecture Notes in Mathematics 757. Springer-Verlag, Berlin.
- Hart, J.D. (1991) "Kernel Regression with Time Series Errors," (to appear in the *Journal of the Royal Statistical Society Series B*)
- IMSL Incorporated (1984) *IMSL Library Reference Library*, Houston.
- Priestley, M.B. and Chao, M.T. (1972) "Non-parametric function fitting," *Journal of the Royal Statistical Society Series B* 34 385-392.

Silverman, B.W. (1984) "Spline Smoothing: The Equivalent Variable Kernel Method," *Annals of Statistics* 12 898-916.

Silverman, B.W. (1985) "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting," *Journal of the Royal Statistical Society Series B* 47 1-52.

Truong, Y.K. (1989) "Nonparametric curve estimation with time series errors," preprint.

Velleman, P.F. (1989) *Data Desk Handbook*, Odesta Corporation, IL.

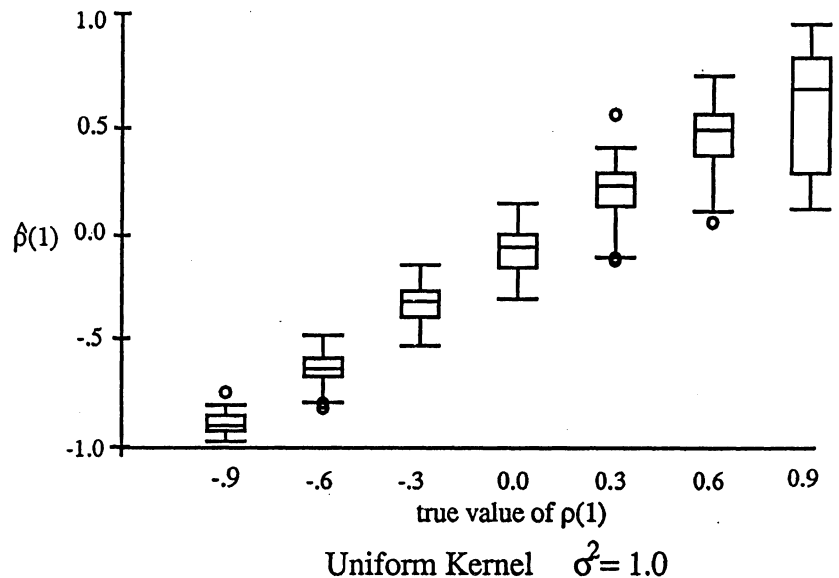
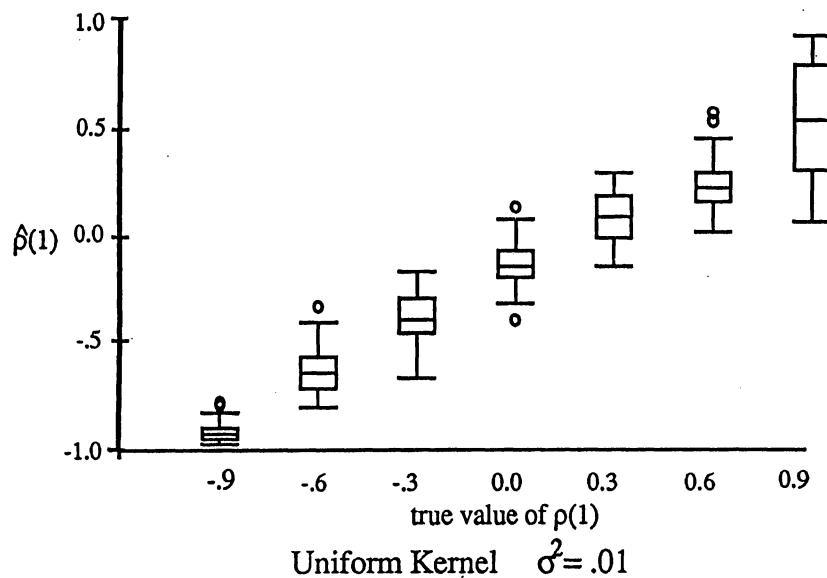
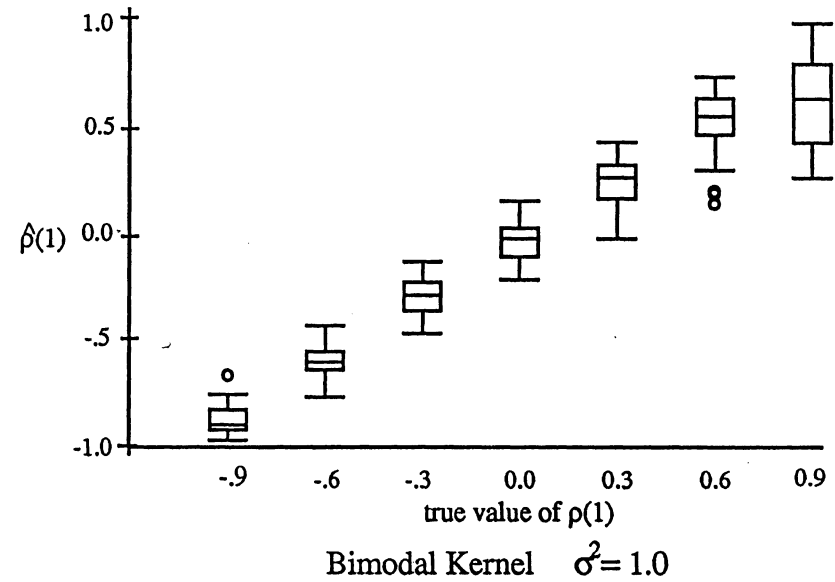
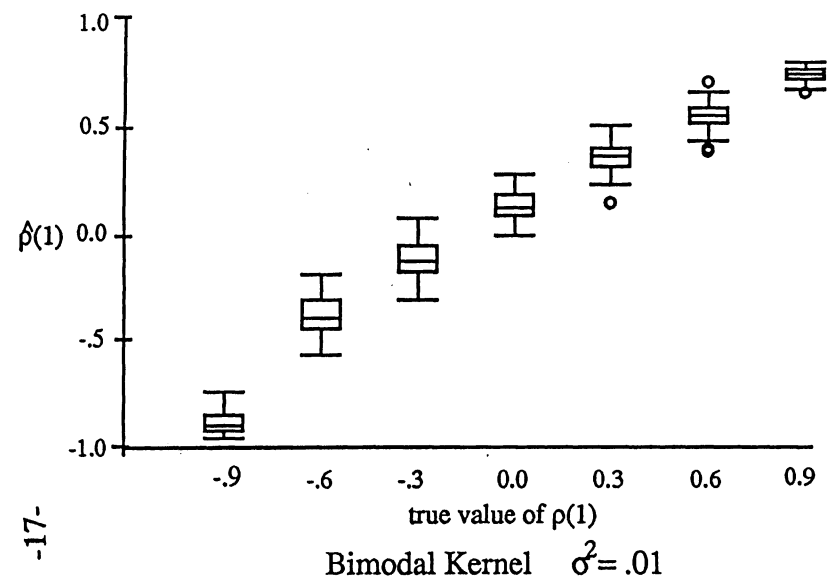


Figure 1: Estimates of $\rho(1)$ using the bimodal and uniform kernels, and various values of σ^2 . Construction of the boxplots is explained in detail in Velleman (1989). Circles denote points .15 to 3 box lengths from the nearest edge. Asterisks denote points more than 3 box lengths from the nearest edge.

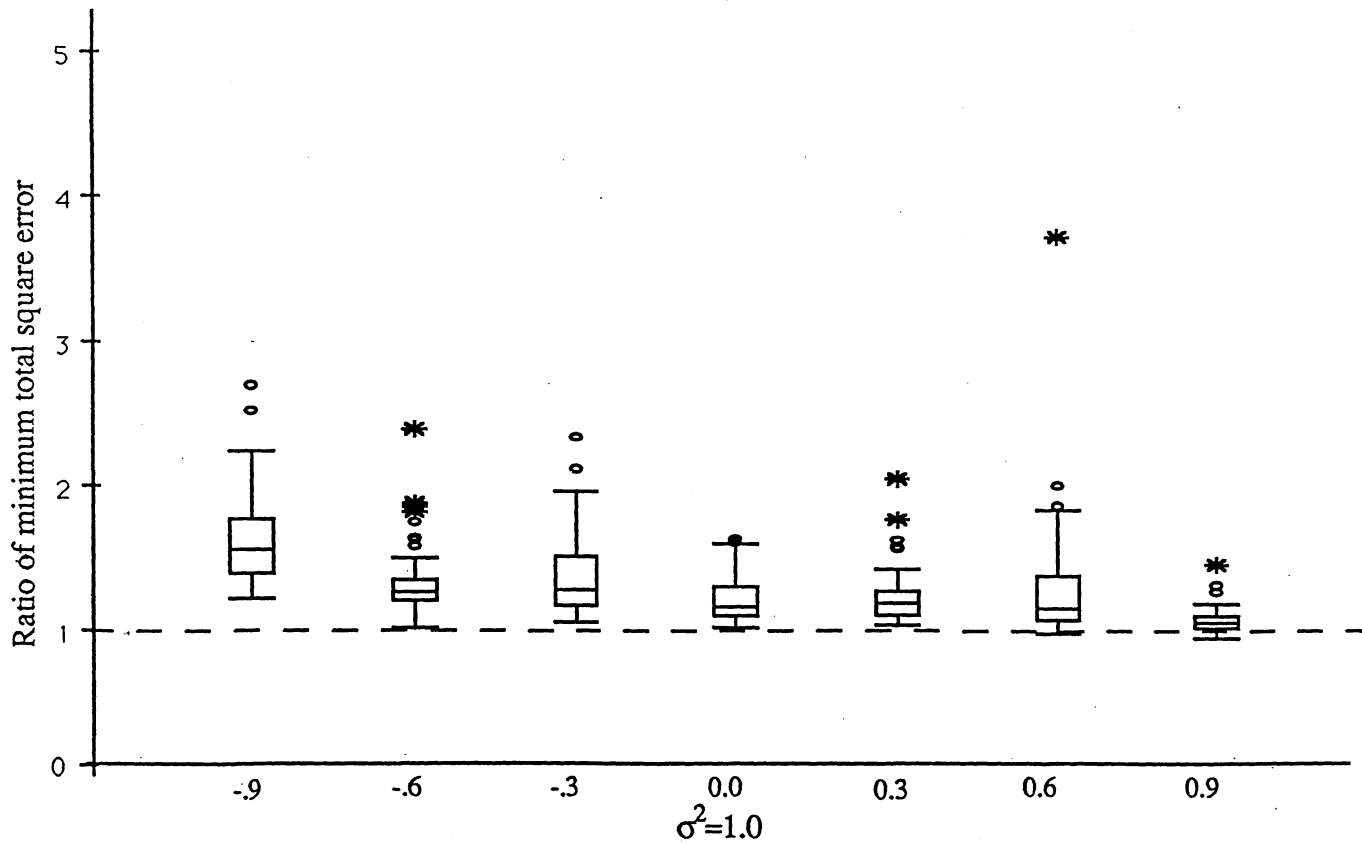
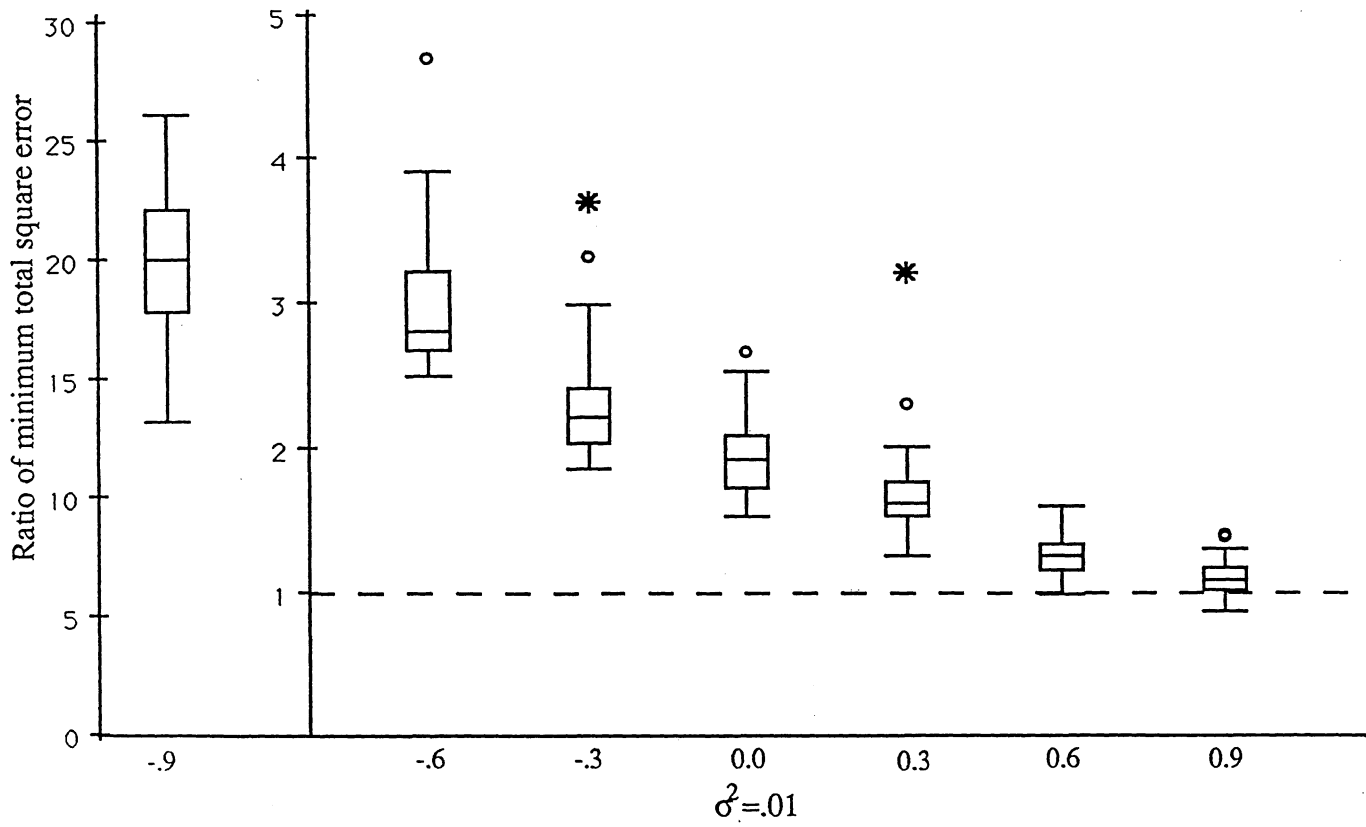


Figure 2: Ratio of the minimum total square errors for estimating the regression function comparing the bimodal kernel to the quadratic kernel .

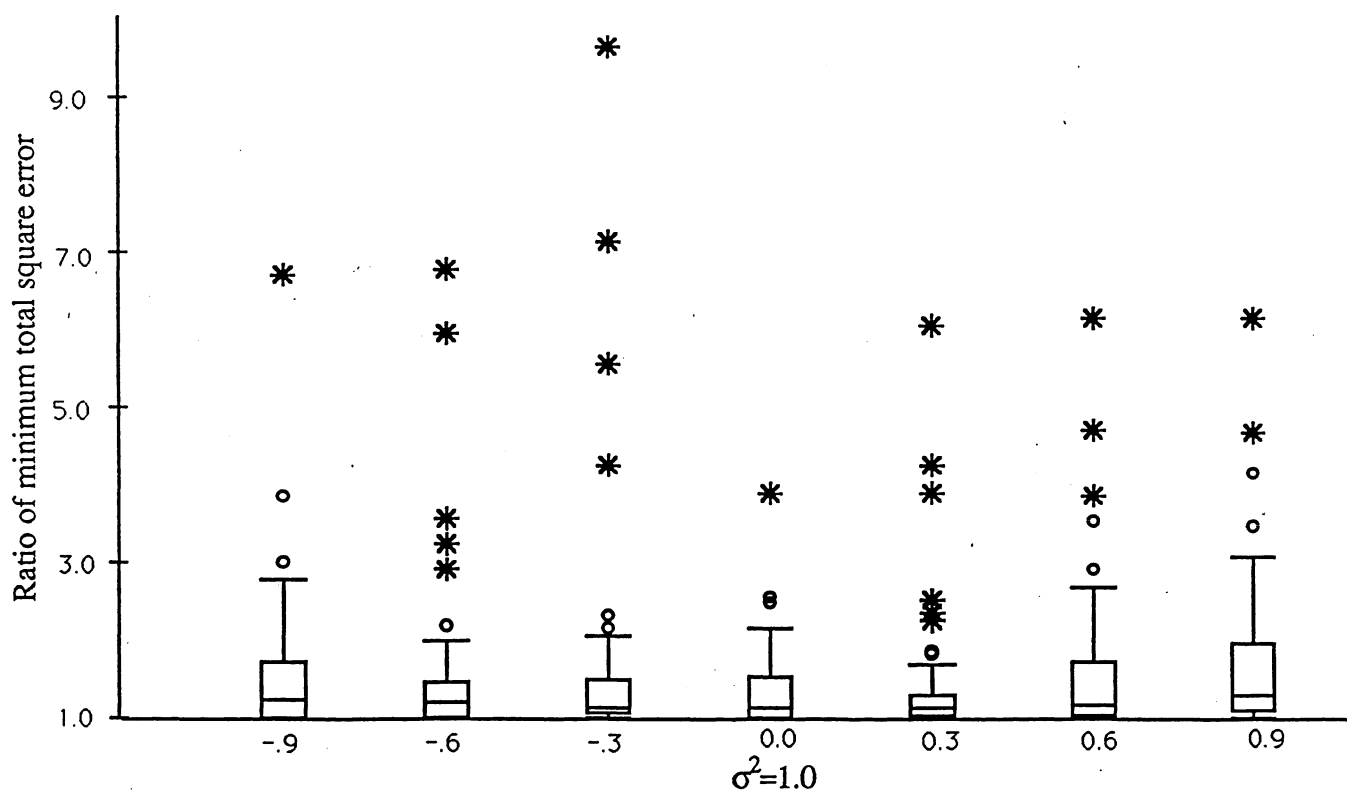
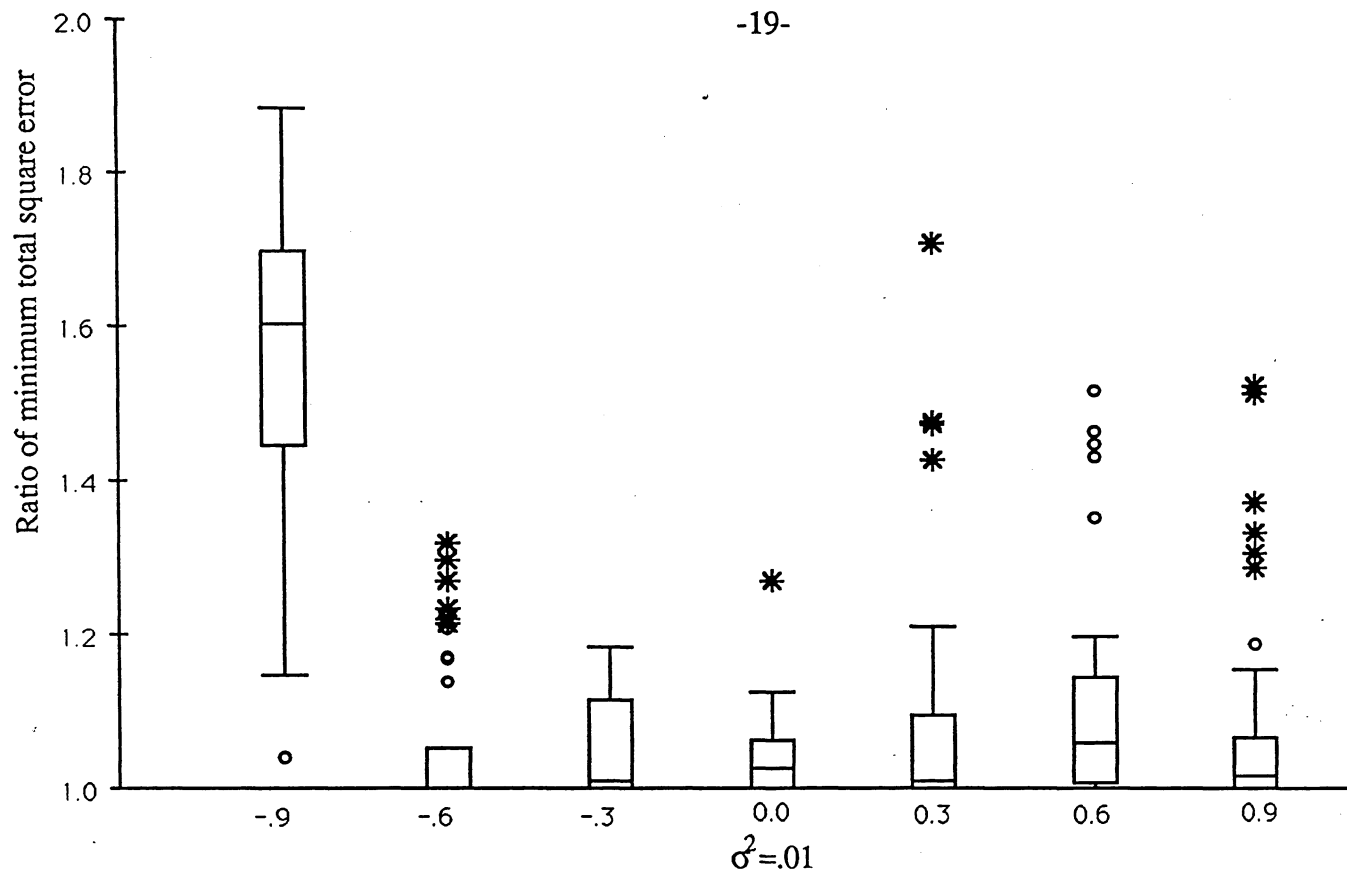


Figure 3: The ratio of total squared errors of the iterated fit of the regression function to the minimum total square error, when the bimodal kernel was used to compute the correlation function and the quadratic kernel was used to compute the regression function.

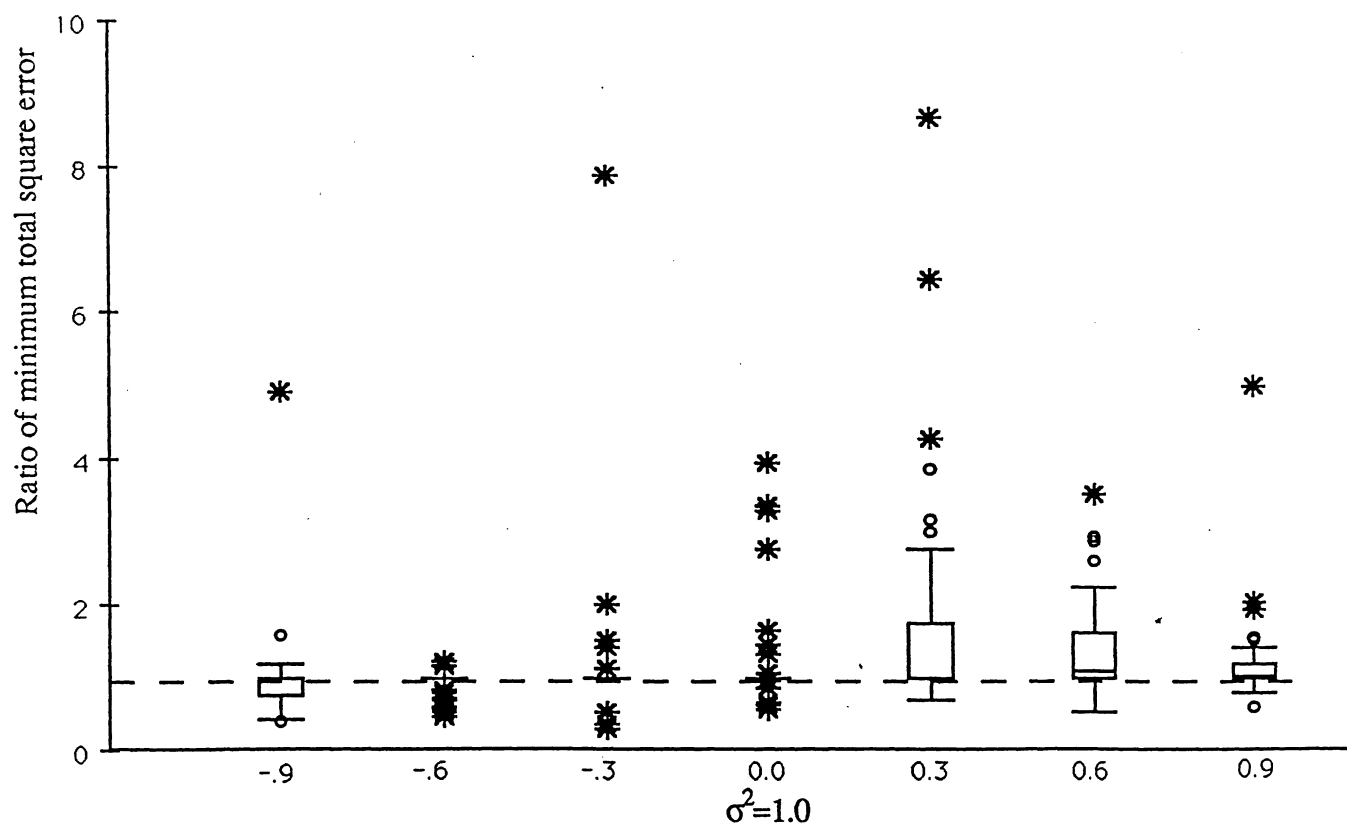
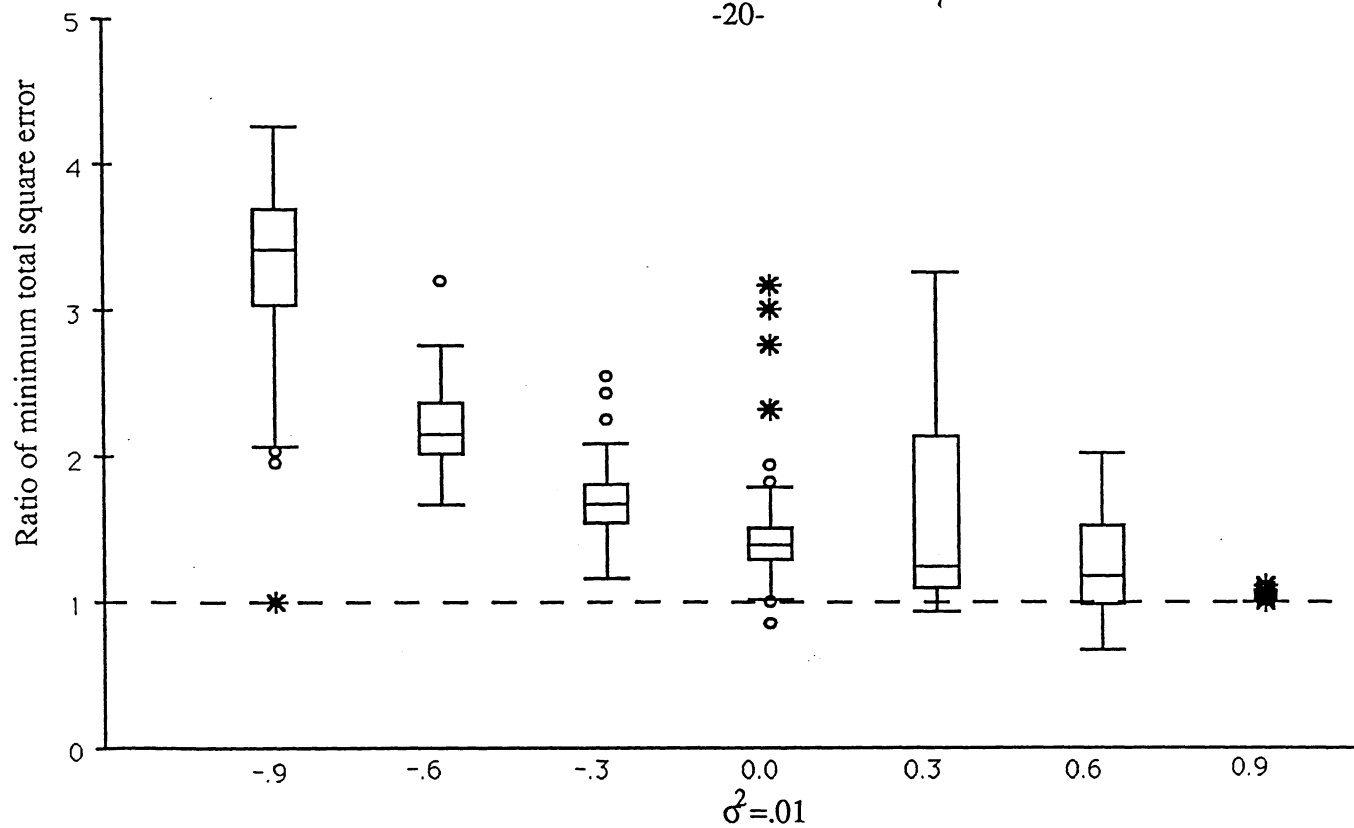


Figure 4: Ratio of total squared errors of the fully iterative procedure to the two-step procedure. The bimodal kernel was used to estimate the correlation function, and the uniform kernel was used to estimate the regression function, for both procedures.